06311 Abstracts Collection
Sensor Data and Information Fusion in Computer Vision and Medicine
— Dagstuhl Seminar —

Joachim Denzler[1], Joachim Hornegger[2], Jürgen Kittler[3] and Calvin R. Maurer JR.[4]

[1] Univ. Jena, DE
denzler@informatik.uni-jena.de
[2] Univ. Erlangen, DE
Joachim.Hornegger@informatik.uni-erlangen.de
[3] Univ. of Surrey, GB
j.kittler@eim.surrey.ac.uk
[4] Stanford Univ., US
crmaurer@stanford.edu

Abstract. From 30.07.06 to 04.08.06, the Dagstuhl Seminar 06311 "Sensor Data and Information Fusion in Computer Vision and Medicine" in the International Conference and Research Center (IBFI), Schloss Dagstuhl. Sensor data fusion is of increasing importance for many research fields and applications. Multi-modal imaging is routine in medicine, and in robitics it is common to use multi-sensor data fusion. During the seminar, researchers and application experts working in the field of sensor data fusion presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. The second part briefly summarizes the contributions.

Keywords. multi-sensor fusion, multi-modal perception, multiple expert fusion, fusion paradigms, multi-modal and intra-modal experts, non-rigid registration, human robot interaction, attention systems, computer vision, image processing, medical image analysis, multi-modal tissue classification, intensity correction, real-time tracking, non-parmetric density estimation, assignment problem, artificial voice

06311 Executive Summary – Sensor Data and Information Fusion in Computer Vision and Medicine

Today many technical systems are equipped with multiple sensors and information sources, like cameras, ultrasound sensors or web data bases. It is no problem to generate an exorbitantly large amount of data, but it is mostly unsolved how

to take advantage of the expectation that the collected data provide more information than the sum of its parts. The design and analysis of algorithms for sensor data and information acquisition and fusion as well as the usage in a differentiated application field was the major focus of the Seminar held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. 24 researchers, practitioners, and application experts from different areas met to summarize the current state-of-the-art technology in data and information fusion, to discuss current research problems in fusion, and to envision future demands of this challenging research field. The considered application scenarios for data and information fusion were in the fields of computer vision and medicine.

Keywords: sensor and data fusion, adaptive fusion, multimodal fusion, multiple classifier fusion, computer vision, robotics, medical imaging

Joint work of: Denzler, Joachim; Hornegger, Joachim; Kittler, Jürgen; Maurer JR., Calvin R.

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/854

Efficient Fusion of Histograms for Real-Time Tracking Using Mean-Shift and Trust-Region Optimization

Ferid Bajramovic (Universität Jena, D)

Histogram based real-time object tracking methods, like the Mean-Shift tracker of Comaniciu/Meer or the Trust-Region tracker of Liu/Chen, have been presented recently. The main advantage is that a suited histogram allows for very fast and accurate tracking of a moving object even in the case of partial occlusions and for a moving camera.

The problem is which histogram shall be used in which situation. In this paper we extend the framework of histogram based tracking. As a consequence we are able to formulate a tracker that uses a weighted combination of histograms of different features. We compare our approach with two already proposed histogram based trackers for different historgrams on large test sequences availabe to the public. The algorithms run in real-time on standard PC hardware.

Keywords: Real-time tracking, histogram combination

Joint work of: Bajramovic, Ferid; Grässl, Christoph; Denzler, Joachim

Full Paper:
http://herkules.inf.uni-jena.de/biborb/bibs/cv/papers/Bajramovic05:ECO.pdf

See also: Bajramovic, F.; Grässl, C.; Denzler, J.: Efficient Combination of Histograms for Real-Time Tracking Using Mean-Shift and Trust-Region Optimization. Proceedings of the 27th DAGM Symposium on Pattern Recognition, pp 254-261. 2005.

Multi Sensor and Information Fusion for Advanced Driver Assistance

Gregory Baratoff (Siemens VDO Automotive AG - Lindau, D)

We discuss automotive advanced driver assistance (ADAS) systems such as lane departure warning and adaptive cruise control, and in particular the different sensors they use to perceive the environment around the car. Depending on the task, these systems extract the location of lanes, of cars and other relevant objects. Based on this information, the system either informs or warns the user of dangerous situations, e.g. an impending lane departure, or takes over control to keep a safe distance to the preceeding vehicle.

The measurements provided by these sensors must be very reliable because they are supposed to provide safe assistance when the driver is distracted or inattentive and might not react in time to avoid an accident. In addition, market pressure to offer such systems at affordable prices translates into the need to offer low-cost and low-footprint solutions. For these reasons, it is imperative that the best possible use is made of the information provided by these sensors.

In this talk, we present two systems for vehicle following. The first one is a purely vision-based system for detection and tracking of vehicles. Based on an information fusion approach, it uses multiple cues extracted from the image sequence such as texture, edges, and motion, to initially detect a vehicle in the own lane and to then track it over long periods of time. A shortcoming of this system is that the estimate of the distance of the preceding vehicle - which is the main parameter used in an adaptive cruise control system - is not very reliable. The second system extends the first one by adding a radar and a lidar sensor. These additional sensors provide much improved target distance (lidar) and velocity (radar) estimates. We present our multisensor fusion approach based on an extended Kalman filter which fuses these complementary information sources in the target vehicle state space in a robust way and enables a safe longitudinal control required in an adaptive cruise control system.

Keywords: Multi-sensor fusion vision radar lidar automotive EKF

Joint work of: Baratoff, Gregory; Feiten, Wendelin

On Fusion of Multiple Views for Active Object Recognition

Joachim Denzler (Universität Jena, D)

In this talk we present an approach to solve the problem of choosing optimal views (viewpoint selection) and the fusion of these for an optimal 3D object recognition (viewpoint fusion). We formally define the selection of additional views as an optimization problem and we show how to use reinforcement learning for viewpoint training and selection in continuous state spaces without user

interaction. We also present an approach for the fusion of multiple views based on recursive density propagation.

The experimental results show that our viewpoint selection is able to select a minimal number of views and perform an optimal object recognition with respect to the classification.

A Multi-modal Attention System for Robotic Assistants

Gernot Fink (Universität Dortmund, D)

Today's robots more and more lay off their industrial roots and enter areas of daily human life, e.g. as service or entertainment robots. When building such systems many challenging problems have to be solved in order to enable a successful communication with humans. In addition to the robust perception of e.g. faces or spoken language the control of the robot's attention is of fundamental importance. For a robotic assistant ättentioncomprises not only the selection and interpretation of perception results but also the appropriate active control of the sensors and the generation of feedback to human interaction partners.

In this talk I will present research performed when I was still at Bielefeld University together with colleagues from the Applied Computer Science Group. I will describe the multi-modal attention system developed for the mobile robot BIRON - the "Bielefeld Robot Companion". The system is based on methods for visual and auditory perception of persons. The detection results generated for the individual modalities are combined to form multi-modal perception hypotheses. The modality integration scheme applied - the so-called multi-modal anchoring - additionally enables the robust tracking of multiple person hypotheses over time. Furthermore, from the visual and acoustic data gathered it can be decided whether persons are likely to address the robot by spoken language. Those are considered potential communication partners. An interaction with the robot can be initiated by a greeting phrase. The robot's attention will then remain focussed on the current communication partner throughout the following dialog. Feedback about the current focus of attention is provided by the orientation of the active camera and the robot's base.

The capabilities of the attention system will be exemplified by videos from interactions with BIRON.

Keywords: Multi-modal perception, attention system, human-robot interaction

Mathematical Models and Numerical Methods for Image Registration

Bernd Fischer (Universität Lübeck, D)

In this talk we present a general and unified approach to image registration. It covers central problems arising in typical applications.

The main focus is on the derivation of a mathematical model, on state of the art numerical schemes, and on their sound implementation.

Joint work of: Fischer, Bernd; Modersitzki, Jan

Priors for Nonrigid Multimodal Image Registration in Computer Vision and Medicine

Christoph Gütter (Siemens - Princeton, USA)

The introduction of prior knowledge has greatly enhanced numerous purely low-level driven image processing algorithms. In this work, we focus on the problem of nonrigid image registration. A number of powerful registration criteria have been developed in the last decade, most prominently the criterion of maximum mutual information.

Although this criterion provides for good registration results in many applications, it remains a purely low-level criterion. As a consequence, registration results will deteriorate once this low-level information is corrupted, due to noise, partial occlusions or missing image structure. In this paper, we will develop a Bayesian framework that allows to impose statistically learned prior knowledge about the joint intensity distribution into image registration methods. The prior is given by a kernel density estimate on the space of joint intensity distributions computed from a representative set of pre-registered image pairs. This nonparametric prior accurately models previously learned intensity relations between various image modalities and slice locations. Experimental results demonstrate that the resulting registration process is more robust to missing low-level information as it favors intensity correspondences statistically consistent with the learned intensity distributions.

Keywords: Non-rigid registration, multi-modal, prior knowledge, non-parametric density estimation

Full Paper:
http://www-cvpr.iai.uni-bonn.de/pub/pub/cremers_et_al_cvpr06.pdf

See also: Cremers, D.; Gütter, C.; Xu, C.: CVPR, volume 2, pp 1777-1783, June 2006.

Incorporating 3D Salient Region Features into a Registration Framework

Dieter Hahn (Universität Erlangen, D)

We present a novel representation of 3D salient region features and its integration into a hybrid rigid-body registration framework. We adopt scale, translation and rotation invariance properties of those intrinsic 3D features to estimate a transform between underlying mono- or multi-modal 3D medical images.

Our method combines advantageous aspects of both feature- and intensity-based approaches and consists of three steps: an automatic extraction of a set of 3D salient region features on each image, a robust estimation of correspondences and their sub-pixel accurate refinement with outliers elimination. We propose a region-growing based approach for the extraction of 3D salient region features, a solution to the problem of feature clustering and a reduction of the correspondence search space complexity. Results of the developed algorithm are presented for both mono- and multi-modal intra-patient 3D image pairs (CT, PET and SPECT) that have been acquired for change detection, tumor localization, and time based intra-person studies. The accuracy of the method is clinically evaluated by a medical expert with an approach that measures the distance between a set of selected corresponding points consisting of both anatomical and functional structures or lesion sites. This demonstrates the robustness of the proposed method to image overlap, missing information and artefacts. We conclude by discussing potential medical applications and possibilities for integration into a non-rigid registration framework.

Keywords: Registration, Saliency, Salient Region Features

Joint work of: Hahn, D. A.; Wolz, G.; Sun, Y.; Hornegger, J.; Sauer, F.; Kuwert, T.; Xu, X.

See also: Hahn, Dieter Arnold; Wolz, Gabriele; Sun, Yiyong; Hornegger, Joachim; Sauer, Frank; Kuwert, Torsten; Xu, Chenyang: A Practical Salient Region Feature Based 3D Multi-Modality Registration Method for Medical Images. Proceedings of SPIE on Medical Imaging, volume 6144, pp 870–879, 2006.

Mumford-Shah Model for One-to-one Contour Matching

Jingfeng Han (Universität Erlangen, D)

The variational Mumford-Shah model has been studied for simultaneous edge detection and edge matching. However, this model is asymmetrical respect to the phase field function and the spatial transform. It prevents the preservation of the images topology during registration. Herein, we present a new symmetrical variational model devoted to solve the problem. In this model two phase field functions and two non-rigid transforms are used. The pair of phase field functions takes charge of the edge detection of two images. Meanwhile, the pair of consistent spatial transforms matches two images in two directions according to the law of bijectivity. The optimization process is guided by a generalized gradient flow to guarantee smooth relaxation. A multi–scale implementation scheme is applied to ensure the efficiency of the algorithm. We are going to report the moment experiment result of two intersubject monomodal registration problems (CT/CT and MRI/MRI).

Keywords: Mumford-Shah model, contour matching, image registration

Joint work of: Han, Jingfeng; Berkels, Benjamin; Rumpf, Martin; Hornegger, Joachim; Droske, Marc; Fried, Michael; Scorzin, Jasmin; Schaller, Carlo

Information fusion in decision making systems

Josef Kittler (University of Surrey, GB)

Any pattern recognition system fuses measurement information to reach a decision about the identity of an object or phenomena to be recognised. When designing such a system the key question is how the available measurement information should be combined in order to find the best possible separation of pattern classes. In this paper we address the problem of information fusion at somewhat different level. We shall consider to what extent the performance of a decision-making system can be enhanced by building multiple experts, in order to combine their opinions to reach a consensus decision. To this end we could draw on experts that base their opinions about an object to be identified either on the same measurement data, or on different sensory modalities as well as on contextual information sources. However, the focus of the paper will be on intramodal and multimodal expert fusion.

It is well recognised that the process of classifier design is detrimentally affected by serious lack of knowledge of the underlying probability distributions of pattern classes. This is manifest in structural and estimation errors which affect the accuracy of the models that are inferred as part of the classifier design process. A body of evidence suggests that Bayesian estimation methods can provide a measure of protection against severe modelling errors and their use results in better pattern recognition system designs, with significantly boosted performance. The essence of Bayesian estimation is to integrate over the probability distributions of the system design parameters. It will be argued that this integration can be accomplished by building and combining multiple classifiers. The focus of the discussion will then be on multimodal fusion, where sensory information from multiple sensors is combined to accumulate complimentary sources of information about the objects to be classified. The problem of intramodal fusion will then be considered. The merit of multiple classifier fusion will be illustrated on problems drawn from different application domains.

Keywords: Multiple Expert Fusion, Fusion Paradigms, Bayesian Estimation, Multimodal and Intramodal Experts

Probabilistic, Adaptive Sensor Data Fusion for 3-D Object Tracking

Olaf Kähler (Universität Jena, D)

Object tracking is a typical application scenario for data fusion algorithms. Different tracking methods for structured and uniform objects can be combined to achieve robust 2-D tracking, different cameras or modalities have to be fused to reconstruct the 3-D position of the object.

We present a framework handling both uni- and multimodal fusion in a common, probabilistic way. Further we investigate the adaptive fusion algorithms

Democratic Integration and STAPLE fusion in the context of this application scenario. Experimental evaluations will conclude the presentation.

Keywords: Object Tracking, Democratic Integration, STAPLE

Optimization in problems of image registration

Günther Leugering (Univ. Erlangen-Nürnberg, D)

We discuss various distance measures and regularization terms in the context of optimization in image registration. We point out that image registration in its variational formulation can alternatively be interpreted in the context of optimal control of partial differential equations. We then discuss, on the discrete level, coupled assignment and registration problems which can be treated by global optimization methods. These discrete models tend, by the way of Γ-convergence, to infinitedimensional optimization problems, which, in turn, are equivalent to the Monge-Kantorovich mass transportation problem. We give algortihms and provide numerical results. Adjoining the Monge-Kantorovich or Wasserstein measure with normal information seems to be a promising direction.

Keywords: Distance measures, assignment problem, mass-transportation, registration

Supervised learning on incomplete multimodal data

Björn H. Menze (Universität Heidelberg, D)

With the increasing dimensionality of multimodal information, the complexity of its analysis increases considerably. Display and evaluation of bi-modal data is feasible by plain overlay, as applied e.g. on PET & CT images in medical diagnostics; up to three features can easily be visualized by false-color maps, routinely used in the display of satellite imagery. The parallel evaluation of more than three modalities, however, becomes more and more complex. Here the design of machine-based, multivariate decision rules allows to cope with the problems arising from the dimensionality of such data. Unfortunately, an increasing number of (independent) sensors also increases the overall chance of a failure of one of them, leading to missing values in the feature vector. On high dimensional multimodal data, this incompleteness – or missingness – becomes a major concern in the design of the decision rule.

In this paper different approaches from statistical learning will be reviewed which are able to cope with incomplete data both in training and testing: imputation strategies, naive bayes, decision trees, extensions to discriminant analysis, and nonparametric models.

It will be discussed how these algorithms can be benchmarked on incomplete data sets.

Keywords: Missing data, medical image analysis, multimodal tissue classification

A Unified Approach to Registration and Intensity Correction (RIC)

Jan Modersitzki (Universität Lübeck, D)

In many medical applications, registration and intensity correction problems do intertwine. We propose a new methodological framework for a joint registration and intensity correction (RIC). This framework is based on minimization of a joint energy J with various degrees of freedom. The essential advantages of the RIC approach are: combination of registration and intensity correction into one unified framework and thus joint minimization. The performance of the RIC approach is demonstrated on a variety of applications ranging from MRI to histology.

Keywords: Optimization, image processing, registration, matching, intensity correction

Joint work of: Modersitzki, Jan; Papenberg, Nils

Local correlation for improving 3D-pointcloud fusion

Christoph Munkelt (Fraunhofer Institut - Jena, D)

Whole body 3D-measurement depends on multiple cameras taking multiple views of the object to be scanned. While phasogrammetry handles the general data fusion reliably with high precision, analysis of local correlation can improve 3D-pointcloud fusion even further. Usage of a volumetric data model enables spatial evaluation of point- / patch neighbourship as well as associated criteria.

Local coordinate adjustments reduce inter-patch gapes caused by erroneous bundle block adjustments. Analysis of regional planarity across patches helps to identify and eliminate outliers. Quality based patch selection ensures consistency in patch overlap regions. It also enables controlling of local point density for the purpose of a homogeneous final point cloud.

Keywords: 3D-pointcloud fusion, spatial locality analysis, noise reduction

Localisation and Tracking of Audiovisual Objects by Multimodal Data Fusion

Rudolf Rabenstein (Universität Erlangen-Nürnberg, D)

Audiovsiual objects can be recorded either through their visual appearance with cameras or through their emitted sound waves with microphones. Both modalities are suitable for the localisation of stationary sources or for tracking of moving sources.

However, also both modalities are subject to specific impairments. Image sequences may suffer from occlusions or lighting changes while microphone recordings may contain reverberation or background noise. These impairments of localisation and tracking results may be reduced when both modilties are combined with a suitable method for multimodal data fusion. Appropriate features for this purpose are the position and the velocity of each object. Considering these features as the state of the object, their estimation for each modality and the fusion to a multimodal result can be formulated as a recursive state estimation. It can be realised with the concepts of Kalman filtering or particle filtering. Examples show that multimodal tracking methods outperform single sensor methods in terms of tracking accuracy.

Keywords: Localisation, Tracking, Audiovisual, Multimodal, Data Fusion, Kalman Filter, Particle Filter

Evolution of Information Fusion in Computer Vision and Medicine

Daniel Russakoff (Stanford University, USA)

Though similar fields in general, the worlds of computer vision and medical image analysis are different in a number of ways. Each serves different constituencies and, as a result, each has different design constraints and preferred methodologies. In particular, this divergence can be seen in the work on information fusion. Whereas in computer vision, the majority of problems are traditionally motivated by robotics and surveillance applications, most medical image analysis problems come from medicine and biology. Though no less important, the information fusion problems posed by the robotics community tend to be much more difficult than those from medicine and biology as they are typically dynamic in nature, must work quickly, and must effectively deal with rapidly changing and unknown environments. By comparison, the majority of information fusion problems in medicine and biology are static and operate under more controlled conditions. Not surprisingly, techniques to perform information fusion have evolved differently in these communities with limited cross-over.

This talk will discuss the problem of information fusion in general and will detail, from a computer vision perspective, some of the successes and trends in the field. Some attention will also be devoted to trends in information fusion in the medical imaging community as well as signs that the two communities are beginning to converge.

Keywords: Information fusion, computer vision, medical imaging

Mixed-Integer Programming Formulation of the Registration Problem with an Application in Phoniatrics

Michael Stiglmayr (Univ. Erlangen-Nürnberg, D)

Point based registration algorithms consist of two parts: Finding an assignment of automatically (or manually) chosen points (landmarks) and finding a transformation such that the "distance" of assigned points is minimal. These two coupled problems can be considered as a mixed integer optimization problem. The assignment variables are integer (binary: two points are assigned or not assigned) and the transformation variables can be optimized in a real vector space.

This formulation gives the rare opportunity to find a global optimal solution for a registration problem. We will present first academic results on globally optimized registration problems as well as open questions in the field of computability of these complex problems.

An application to the modeling of the voice restoration after total excision of the larynx is also discussed.

Keywords: Point-based Registration, Mixed-Integer Optimization, Phoniatrics, Artificial Voice

Joint work of: Stiglmayr, Michael; Schwarz, Raphael; Klamroth, Kathrin; Leugering, Günter; Lohscheller, Jörg

Information fusion in vision-based human-machine interaction systems

Sven Wachsmuth (Universität Bielefeld, D)

In the last years, computer vision systems have gained more and more successes in real world applications. However, most of these systems do not solve the generic vision problem. Instead they combine very specific vision behaviours. For example in the field of human-machine interaction systems typically need to combine object recognition, action recognition, pose localisation, scene classification, and even speech understanding.

All available information has to be fused in order select an appropriate system response. In my talk I will tackle a couple of problems that arise in this type of application. What are generic representations for fusing different types of specific information? How can we deal with asynchronously operating behaviors that have there own local control strategy?

Joint work with: Sebastian Wrede, Marc Hanheide, Gerhard Sagerer, Mike Jamieson, Sven Dickinson, Suzanne Stevenson How can we automatically find an appropriate vocabulary for mapping information from different modalities? What are the benefits of a centralized fusion strategy?

The presentation will sketch work conducted in the VAMPIRE project and in the area of integrating vision and language. The first part will focus on an

active memory infrastructure (AMI) that provides a generic framework for fusing structured data. Visual behaviors interact asynchronously via the AMI in a data-driven fashion. Synchronization between behaviours is achieved by petri-net structures that are linked to memory events.

The second part of the talk focuses on the problem of integrating language and vision. A visual vocabulary that can be linked to words by a probabilistic translation model is learned from a captioned image dataset.

Joint work of: Wrede, Sebastian; Hanheide, Marc; Sagerer, Gerhard; Jamieson, Mike; Dickinson, Sven; Stevenson, Suzanne

Keywords: Computer vision systems, human-machine interaction, fusion of multi-modal information